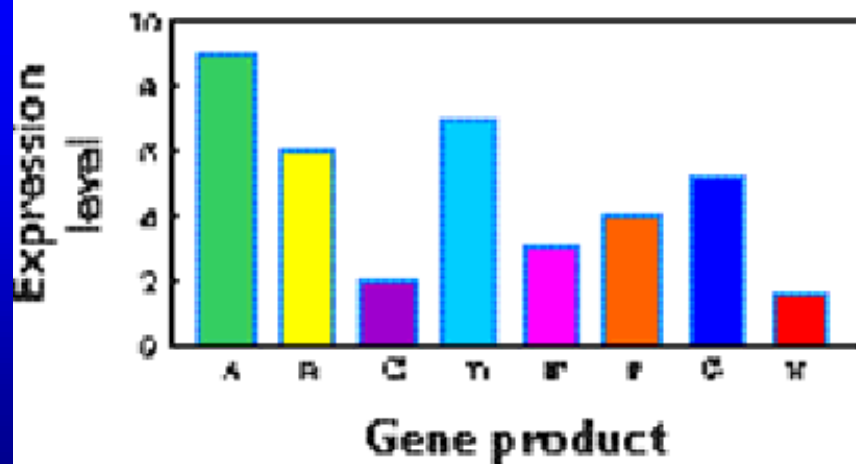


Overview-SAGE

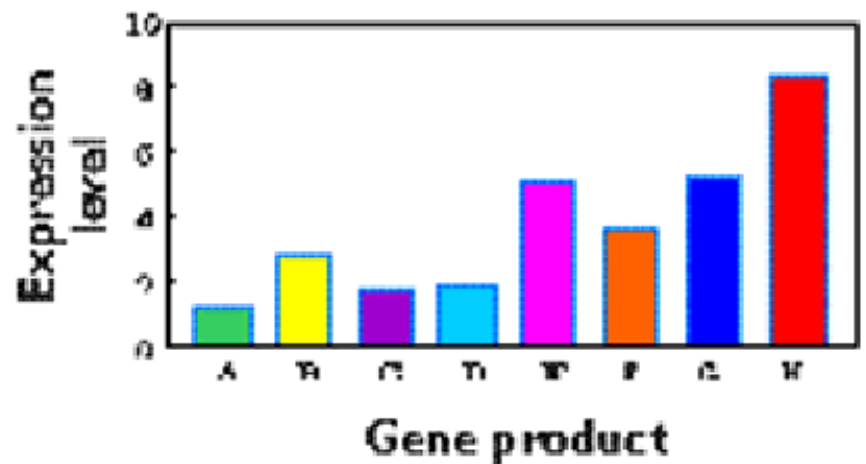
- Principles
- Methodological outline
- Analytic strategies
- Strengths and limitations
- Examples of applications

SAGE: Serial Analysis of Gene Expression

- Simultaneous and quantitative comparison of gene-specific sequence tags.



Normal

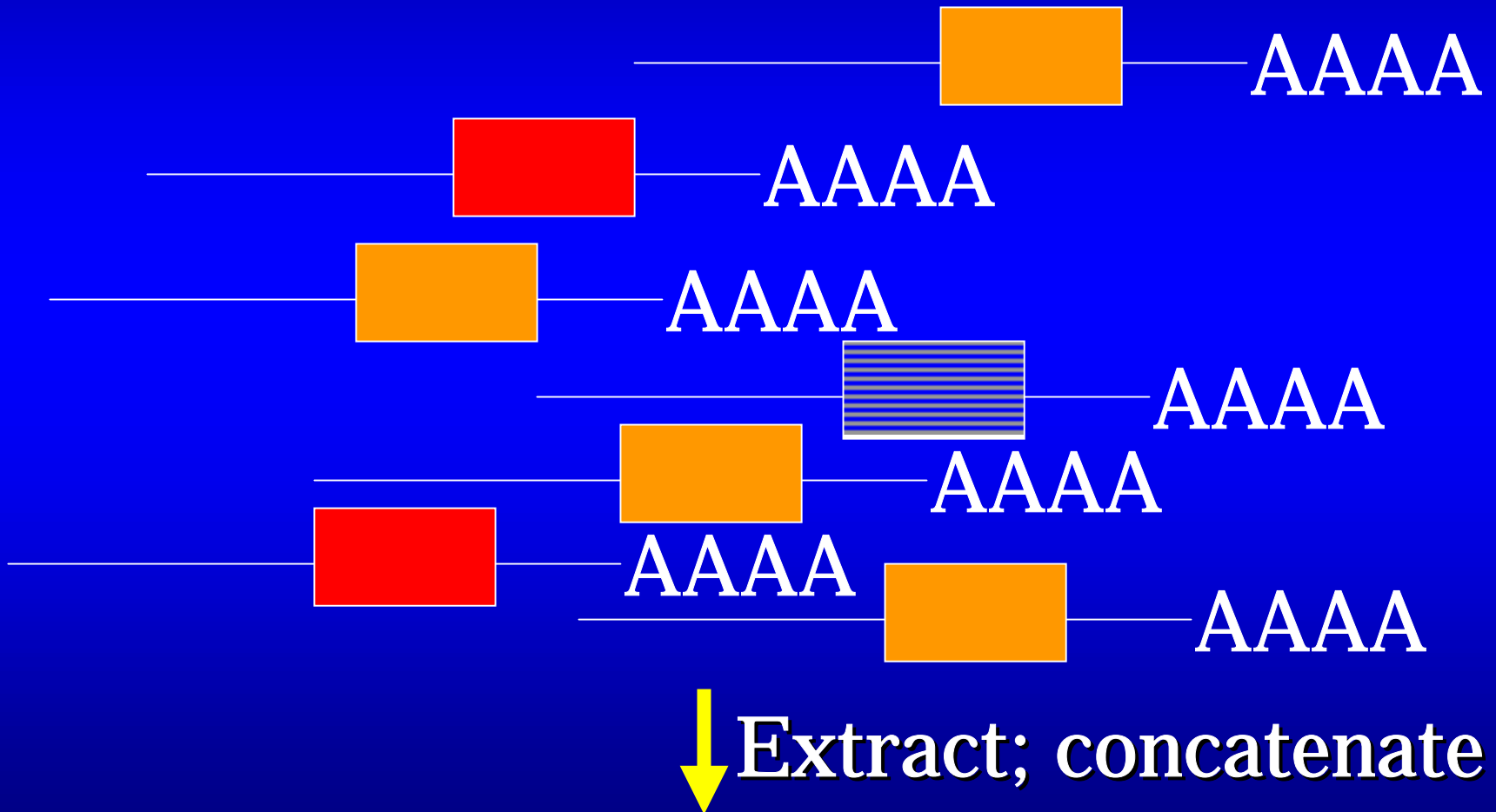


Disease

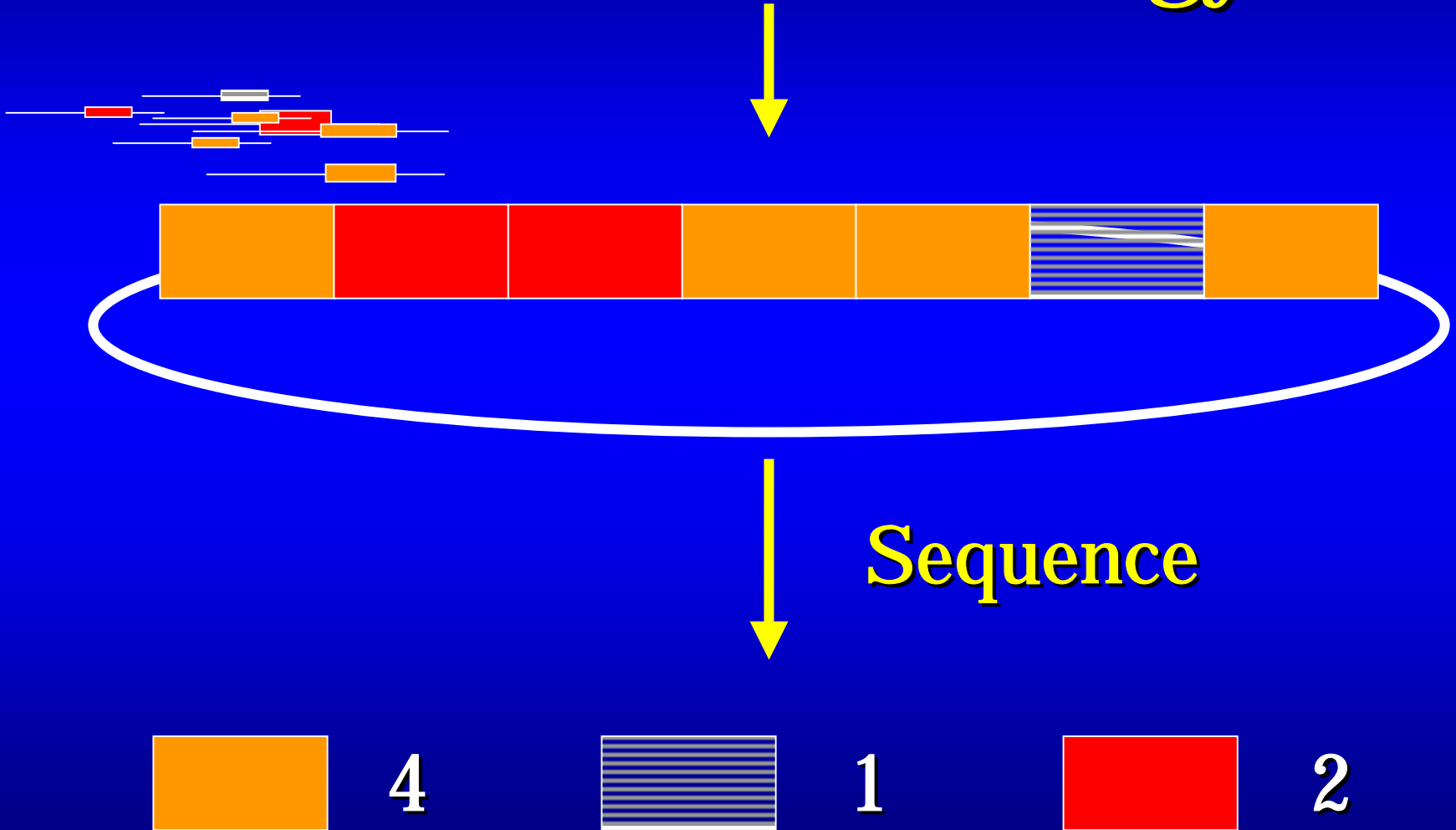
SAGE: Principles

- No prior knowledge of genes expressed (The hypothesis is that there is no hypothesis).
- mRNAs represented by short (9 to 13 bp) sequence tags
- Concatenation of the tags for efficient sequencing
- Amplification structured to avoid distortion

SAGE: Methodology



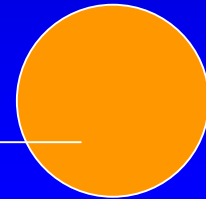
SAGE: Methodology



SAGE: Methodology

TE

A GGGACATGNNNAAA
CCCTGTACNNNTTTT



AE

A GGGACATGNNNNNNNNCATGCC **B**
CCCTGTACNNNNNNNNGTACGGG

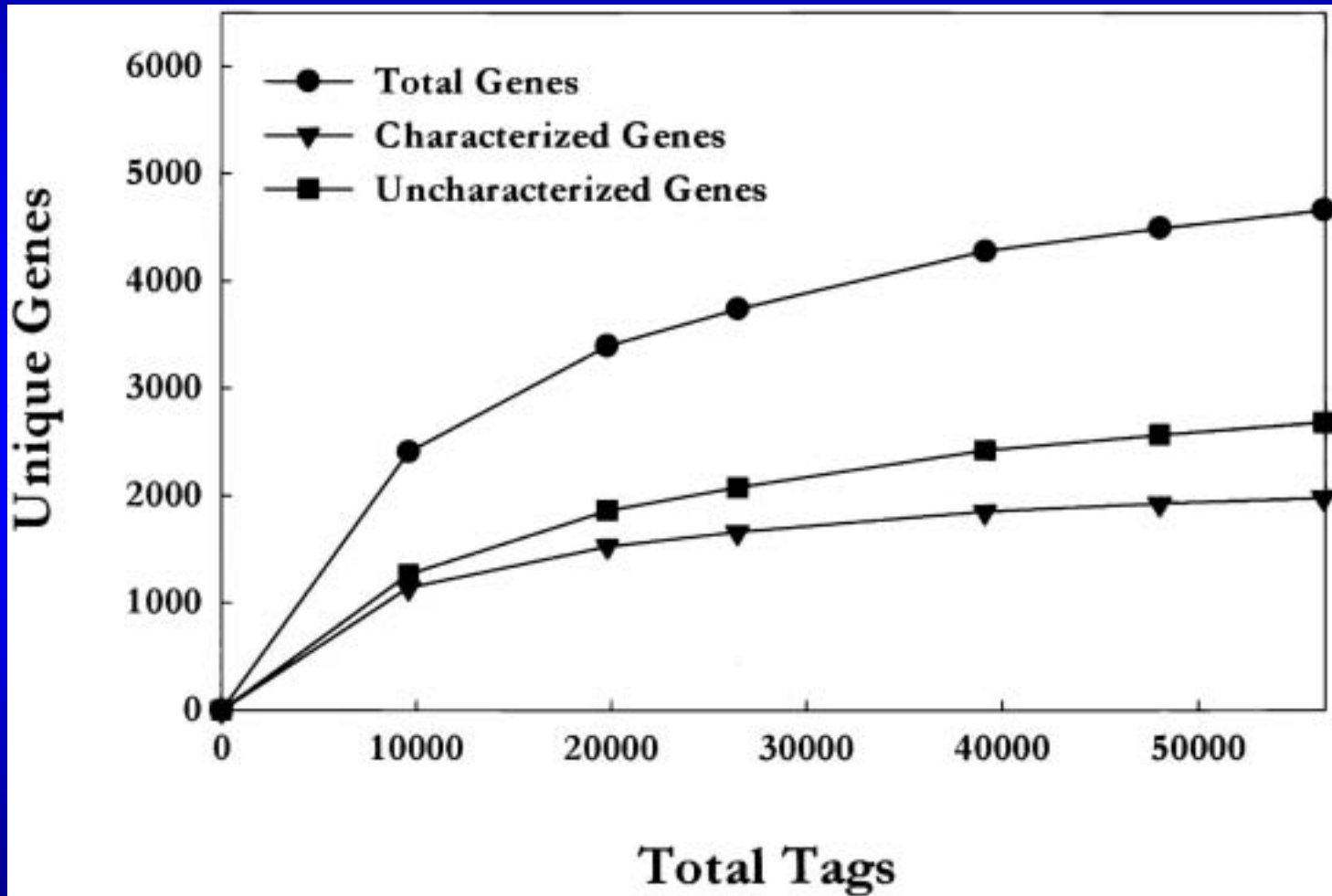


NNNNNNNNCATG
GTACNNNNNNNN

SAGE: Data output

SAGE Tag	Exp1	Exp2
ATAATACATA	249	469
AAAAAAAAAAAA	43	128
AGGACAAATA	144	25
ATACTGACAT	502	612
GCTGCCCTCC	1138	670
CTATCCTCTC	613	707
AGGAGGACTT	180	63
AGCAATTCAA	318	146
ATAATACAAA	10	46
TAGATATAGG	94	22
CAAACCTCCA	68	12
GAAAAAAAAAAA	26	60

How many tags?



Cell, 1997

Genes in Normal Kidney

- **GACTTCACGCC**
Mouse kidney androgen-regulated protein
- **CTATTCCTCTCA**
Plasma glutathione peroxidase
- **TGTAGCCTCAT**
 Na^+/K^+ ATPase γ chain
- **GGCCTTACTTC**
 Na^+/P_i transporter

What makes a tag reliable?

- Tag should be 3'-adjacent to the 3'-most NlaIII (CATG) site
- Matches a well characterized cDNA. ESTs with a polyA signal or tail are next most reliable

CONFIRMATION CRITICAL

SAGE Tag to Gene Mapping

SAGETag (10 bases):



Reliable UniGene clusters matched to this tag:

[Mm.7156](#) : **glutathione peroxidase 3**

SAGE library data for this tag:

No SAGE data for this species and/or anchor.

Summary of UniGene clusters found for this tag:

<i>UniGene cluster id(s)</i>	<i>UniGene cluster title</i>	<i>Tag->cid frequency</i>	<i>Get seqs</i>
	cDNA: well-characterized		
Mm.7156	glutathione peroxidase 3	3/3	
	EST: 3' oriented, 3' label		
Mm.7156	glutathione peroxidase 3	24/24	
	EST: 3' oriented, no label		

SAGetag (10 bases):

GTGGCTCACA

Mus musculus



NlaIII

Reliable UniGene clusters matched to this tag:

[Mm.100791](#) : RIKEN cDNA 2700038G22 gene

[Mm.10098](#) : non-selective cation channel 1

[Mm.103857](#) : RIKEN cDNA 4930556A12 gene

[Mm.1114](#) : galactosidase, alpha

[Mm.117043](#) : transcription termination factor 1

[Mm.11756](#) : chemokine (C-X-C) receptor 2

[Mm.12322](#) : RIKEN cDNA 1110004B06 gene

[Mm.1378](#) : cannabinoid receptor 2 (macrophage)

[Mm.143758](#) : similar to RAP1 protein

[Mm.153276](#) : Mus musculus, clone MGC:8197, mRNA, complete cds

[Mm.156851](#) : RIKEN cDNA 2310050N03 gene

[Mm.158192](#) : RIKEN cDNA C330014O21 gene

[Mm.158221](#) : RIKEN cDNA 1110035E04 gene

[Mm.159789](#) : RIKEN cDNA 2900022B07 gene

+ 20 other clusters

SAGE: Data analysis

Resource

Contact Information

SAGE300	www.sagenet.org
SAGEmap	www.ncbi.nih.gov/SAGE
P value	igs-server.cnrs-mrs.fr/
P value	Email: j.m.ruijter@amc.uva.nl
eSage	Email: ehm@umich.edu
USAGE	http://www.cmbi.kun.nl/usage/bin/login.cgi

Examples

Expression profile from normal mouse kidney

- Analyses of 3,868 tags yielded 1,453 unique kidney transcripts:
 - 42% had known function
 - 35% ESTs
 - 23% were unknown
- Genes that regulate normal renal physiology accounted for 19% of all tag sequences.
- Transcripts encoding proteins which limit tissue injury were expressed.

Comparison of mRNA frequencies

Comparison of mRNA expression profiles from kidneys of normal mice and mice with kidney disease will identify chronic renal disease genes.

SAGE: Results

- 49,588 tags comprising 20,594 unique genes in the kidney transcriptome
- Over 5,000 genes expressed at a level of ≥ 2 copies
- 5,521 not in public databases

SAGE: Statistical analysis

- Differential expression:
 - * 238 genes differentially expressed at $p < 0.05$
 - * 63 genes differentially expressed at $p < 0.01$. Of these 25 underexpressed in kidneys of ROP-Os/+ mice

SAGE: Gene networks

Pubgene, Nature Genetics, 2001;
www.pubgene.uio.no

SAGE: Summary

- Advantages

Comprehensive

Quantitative

Can be done in a small lab

Modifications for small samples,
to reduce frequency of
nonunique tags

SAGE: Summary

- Drawbacks

Labor-intensive

Replication

Analytic tools limited

Need to identify genes encoding tags;
unknowns a particular problem

Is a tag sequence unique for its gene?

Sampling error

Colleagues

- Ashraf El-Meanawy
- Jeff Schelling
- Sudha Iyengar
- Shri Barathan
- Katrina Goddard